Crystal structure of human arginase I complexed with thiosemicarbazide reveals an unusual thiocarbonyl μ-sulfide ligand in the binuclear manganese cluster

Luigi Di Costanzo, Michael E. Pique and David Christianson

SUPPORTING INFORMATION

Materials and methods

Crystal structure determination of unliganded human arginase I

Human arginase I was overexpressed in *Escherichia coli* and purified as previously described.[1] Unliganded human arginase I was crystallized by the sitting drop vapor diffusion method at 21 °C. Drops containing 4.0 μL of enzyme solution [3.5 mg/mL human arginase I, 0.1 mM $MnCl_2$, 1.4 mM methylthiosemicarbazide, 50.0 mM bicine (pH 8.5)] and 2.5 μL of precipitant buffer [0.1 M Hepes (pH 7.0), 12-22% (wt/vol) Jeffamine ED-2001®] were equilibrated against a 1 mL reservoir of precipitant buffer. Rod-like crystals appeared in approximately one week and grew to typical dimensions of 0.3 mm x 0.3 mm x 0.7 mm. Crystals were cryoprotected in the same mother liquor augmented with Jeffamine ED-2001® to a final concentration of 32% (w/v). These crystals yielded diffraction data to 1.90 Å resolution at the Brookhaven National Laboratory on beamline X12B (λ=1.000 Å, 100 K). Data

were reduced with Mosflm[2] and scaled with SCALA from CCP4[3] (Table S1). Diffraction intensities measured from these crystals exhibited symmetry consistent with the apparent space group *P*6 (unit cell parameters a = b = 90.8 Å, c = 69.7 Å). Deviations from ideal Wilson statistics were observed with $\langle I^2 \rangle / \langle I \rangle^2 = 1.5$, indicating the presence of perfect hemihedral twinning.[4] Hence, the true crystallographic space group was *P*3, with 2 arginase monomers from 2 separate trimers in the asymmetric unit. The X-ray crystallographic analysis of human arginase I in twinned crystals has been described previously.[5]

The structure was solved by molecular replacement using the program *Phaser*[6] with chain A of the human arginase I-ABH complex (PDB accession code 2AEB, less inhibitor atoms and water molecules[5]) used as a search probe against twinned data. In order to calculate electron density maps, structure factors amplitudes ($|F_{obs}|$) derived from twinned data ($|I_{obs}|$) were deconvoluted into structure factor amplitudes corresponding to twin domains A and B ($|F_{obs/A}|$ and $|F_{obs/B}|$, respectively) using the structure-based algorithm of Redinbo and Yeates[7] implemented in CNS.[8] Iterative cycles of model building with the graphics program O[9] and refinement using torsion angle dynamics as implemented in CNS improved the protein structure as monitored by R_{twin} and $R_{free/twin}$. Strict noncrystallographic symmetry restraints were used during refinement and released with the appropriate weighting scheme as judged by $R_{free/twin}$ and the residual Fourier map as refinement progressed. No evidence for methylthiosemicarbazide binding was apparent in electron density maps of the active site, nor was there any ambiguity in the interpretation of active site solvent structure. Therefore, we concluded that this structure represented the first structure of native, unliganded human arginase I. Group B-factors were utilized during refinement. In the final stages of refinement the majority of water molecules were automatically fit into residual electron density peaks using a cutoff of 3.0σ, which improved the R_{twin} and $R_{free/twin}$ values to those reported in Table S1. A total of 260 water molecules were included in the refinement. Disordered segments at the N- and C-termini (M1-K4 and N319-K322) were absent in the

experimental electron density and are omitted from the final model.

Crystal structure determination of the human arginase I-thiosemicarbazide complex

The human arginase I-thiosemicarbazide complex was crystallized by the sitting drop vapor diffusion method at 21 °C as described above with the exception that the protein solution contained 1.4 mM thiosemicarbazide instead of methylthiosemicarbazide. Crystals of human arginase I-thiosemicarbazide complex were subsequently soaked in a precipitant solution augmented with 10.0 mM thiosemicarbazide. Crystals were cryoprotected in the same mother liquor augmented with Jeffamine ED-2001® to a final concentration of 32% (w/v). These crystals yielded diffraction data to 1.95 Å resolution at the Brookhaven National Laboratory on beamline X12B (λ=1.000 Å, 100 K). Data were reduced with HKL2000.[10] Diffraction intensities measured from these crystals exhibited symmetry consistent with the apparent space group P*6* (unit cell parameters a = b = 90.6 Å, c = 69.6 Å). As with crystals of unliganded human arginase I, deviations from ideal Wilson statistics were observed with $\langle I^2\rangle/\langle I\rangle^2$ = 1.5, indicating the presence of perfect hemihedral twinning that convoluted the true crystallographic space group of *P*3.

A similar approach to that described above for the structure solution and refinement of the unliganded enzyme structure was employed to solve and refine the crystal structure of the human arginase I-thiosemicarbazide complex. The molecular replacement solution obtained with *Phaser*[6] was very similar to that obtained in the structure determination of the unliganded enzyme. Group B-factors were utilized during refinement. In the later stages of refinement after the majority of water molecules were located, a gradient omit map calculated using routines implemented in CNS[8] clearly revealed a strong peak corresponding to thiosemicarbazide bound in the active site of monomer A in the asymmetric unit. The thiosemicarbazide molecule was refined with full occupancy to an average B-factor of 35 Å^2, consistent

with the average B-factor of 35 $Å^2$ calculated for the entire protein. A total of 239 water molecules were included in later cycles of refinement. Data reduction and refinement statistics are recorded in Table S1. Disordered segments at the N- and C-termini (M1-S5 and N319-K322) were absent in the experimental electron density and are omitted from the final model.

Figures were generated using Pymol.[11]

Cambridge Structural Database (CSD) analysis

The CSD was installed on a Dell Optiplex 170L, and all database manipulations utilized standard CSD software: ConQuest[12] version 5.98, updated to November 2006. The search included the thiosemicarbazide molecule NH_2-NH-C(=S)-NH_2 coordinated to a metal ion with C=S···M^{n+} separations of 1.5 – 3.0 Å. Each molecular fragment retrieved was visually screened to verify the thiosemicarbazide-metal coordination interaction.

Protein Data Bank (PDB) analysis

Experimentally determined protein structures deposited in the Protein Data Bank as of 1 March 2007 (41806 non-bibliographic entries) were searched for "Mn" in ATOM or HETATM records using tools developed for the Metalloprotein Data Base,[13] available from http://metallo.scripps.edu. This search identified 3347 Mn atoms in 959 entries. A 3 Å shell around each Mn was searched for "S" atoms, without incorporating any symmetry operations, yielding 86 Mn-S interactions. Structures having a Mn-sulfur separation longer than 3.03 Å were excluded from further analysis, yielding a total of 77 unique interactions involving Cys Sγ atoms (62), Met Sδ atoms (12), or S atoms of non-protein residues: 9ica.pdb (with STP [2'-deoxyadenosine 5'-O-(1-thiotriphosphate)] S1A, 1.87 Å), 1cjk.pdb (with

phosphothiophosphoric acid-adenylate ester S1G, 2.98 Å), and 2kfn.pdb (with S replacing uracil O3', 2.61 Å). Each individual Mn-S interaction is recorded in Table S2, and histograms of Mn-S coordination distances, angles, and dihedral angles are recorded in Figure S1.

References

1. Mora, A.; del Ara Rangel, M.; Fuentes, J. M.; Soler, G.; Centeno, F. *Biochim. Biophys. Acta* **2000**, *1476*, 181-190.
2. Leslie, A. G. W. In *Crystallography Computing 5. From Chemistry to Biology*, ed.; Moras, D.; Podjarny, A. D.; Thierry, J. C., **1991**; pp 50-61.
3. Collaborative Computational Project, Number 4. *Acta Cryst. D* **1994**, *50*, 760-763.
4. Yeates, T. O. *Methods Enzymol.* **1997**, *276*, 344-358.
5. Di Costanzo, L.; Sabio, G.; Mora, A.; Rodriguez, P. C.; Ochoa, A. C.; Centeno, F.; Christianson, D. W. *Proc. Natl. Acad. Sci.* **2005**, *102*, 13058-13063.
6. McCoy, A. J.; Grosse-Kunstleve. R. W.; Storni, L. C.; Read, R. J. *Acta Cryst. D* **2005**, *61*, 458-464.
7. Redinbo, M. R.; Yeates, T. O. *Acta Cryst. D* **1993**, *49*, 375-380.
8. Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Cryst. D* **1998**, *54*, 905-921.
9. Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. *Acta Cryst. A* **1991**, *47*, 110-119.
10. Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307-326.
11. DeLano, W. L. *The PyMOL Molecular Graphics System*, 2002.
12. Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. *Acta Cryst. B* **2002**, *58*, 389-397.
13. Castagnetto, J. M.; Hennessy, S. W.; Roberts, V. A.; Getzoff, E. D.; Tainer, J. A.; Pique M., E. *Nucleic Acids Res.* **2002**, *30*, 379-382.

Table S1: Data Collection and Refinement Statistics

	Unliganded human arginase I	Human arginase I-thiosemicarbazide complex
Data Collection		
Resolution, Å	39.3 - 1.90	40.0 - 1.95
Total reflections measured[a]	95532 (13018)	59309 (5845)
Unique reflections measured[a]	49544 (7103)	42929 (4507)
R_{merge}[a,b]	0.077 (0.24)	0.146 (0.322)
$I/\sigma(I)$[a]	11.3 (3.3)	6.1 (2.6)
Completeness (%)[a]	97.7 (95.5)	92.0 (97.3)
Multiplicity[a]	1.9 (1.8)	1.6 (1.5)
Refinement		
Reflections used in refinement/test set	42856/2173	37300/1698
R_{twin}[a,c]	0.198 (0.305)	0.169 (0.283)
$R_{twin/free}$[a,c]	0.244 (0.314)	0.219 (0.286)
Protein atoms[d]	4776	4764
Water molecules[d]	260	239
Thiosemicarbazide atoms[d]	--	5
Manganese ions[d]	4	4
R.m.s. deviations		
Bond lengths, Å	0.008	0.006
Bond angles, °	1.5	1.4
Dihedral angles, °	23.4	22.7
Improper dihedral angles, °	1.0	0.9
Average B-factors, Å^2		
Main chain	20	34
Side chain	21	35
Manganese ions	17	28
Thiosemicarbazide	--	35
Solvent	27	32

[a]Number in parentheses refer to the outer 0.1 Å shell of data.
[b]$R_{merge} = \sum|I-\langle I\rangle|/\sum I$, where I is the observed intensity and $\langle I\rangle$ is the average intensity calculated for replicate data.
[c]$R_{twin} = \sum|[|F_{calc/A}|^2+|F_{calc/B}|^2]^{1/2}-F_{obs}|/\sum|F_{obs}|$ for reflections contained in the working set. $|F_{calc/A}|$ and $|F_{calc/B}|$ are the structure factor amplitudes calculated for the separate twin domains A and B, respectively. R_{twin} underestimates the residual error in the model over the two twin-related reflections by a factor of approximately 0.7. The same expression describes $R_{twin/free}$, which was calculated for test set reflections excluded from refinement.
[d]Per asymmetric unit.

Table S2: Protein structures in the PDB with sulfur---Mn^{2+} interactions.

PDB_code	MN ID code	Ligand	S ID code	Distance (Å)
1cjk	MN _ 503	H AGS	801 S1G	2.98
1gg1	MN A 371	A CYS	61 SG	2.79
1gg1	MN B 371	B CYS	61 SG	2.67
1gg1	MN C 371	C CYS	61 SG	2.79
1gg1	MN D 371	D CYS	61 SG	2.64
1kfl	MN A 351	A CYS	61 SG	2.65
1kfl	MN B 2351	B CYS	61 SG	2.80
1kfl	MN C 3351	C CYS	61 SG	2.79
1kfl	MN D 4351	D CYS	61 SG	2.63
1kfl	MN E 5351	E CYS	61 SG	2.80
1kfl	MN F 6351	F CYS	61 SG	2.72
1kfl	MN G 7351	G CYS	61 SG	2.91
1kfl	MN H 8351	H CYS	61 SG	2.69
1n8f	MN A 351	A CYS	61 SG	2.61
1n8f	MN B 351	B CYS	61 SG	2.59
1n8f	MN C 351	C CYS	61 SG	2.61
1n8f	MN D 351	D CYS	61 SG	2.59
1nkh	MN B 807	B MET	344 SD	2.84
1nkh	MN D 808	D MET	344 SD	2.80
1nrh	MN X 700	X CYS	171 SG	2.38
1o0r	MN B 805	B MET	344 SD	2.69
1o0r	MN A 403	A MET	344 SD	2.68
1o23	MN B 807	B MET	344 SD	2.84
1o23	MN D 808	D MET	746 SD	2.66
1oab	MN A 400	A CYS	76 SG	2.61
1oab	MN B 400	B CYS	76 SG	2.63
1of6	MN A 1369	A CYS	76 SG	2.56
1of6	MN B 1369	B CYS	76 SG	2.50
1of6	MN C 1370	C CYS	76 SG	2.48
1of6	MN D 1369	D CYS	76 SG	2.59
1of6	MN E 1369	E CYS	76 SG	2.15
1of6	MN F 1369	F CYS	76 SG	2.49
1of6	MN G 1368	G CYS	76 SG	2.23
1of6	MN H 1368	H CYS	76 SG	2.52
1ofb	MN A 400	A CYS	76 SG	2.68
1ofb	MN B 400	B CYS	76 SG	2.66
1ofq	MN A 1000	A CYS	1076 SG	2.33
1ofq	MN B 2000	B CYS	2076 SG	2.62
1ofq	MN C 3000	C CYS	3076 SG	2.41
1ofq	MN D 4000	D CYS	4076 SG	2.52

1ofr	MN	A	400	A	CYS	76	SG	3.03
1ofr	MN	B	400	H	CYS	76	SG	2.49
1ofr	MN	D	400	D	CYS	76	SG	2.62
1ofr	MN	E	400	E	CYS	76	SG	2.78
1ofr	MN	F	400	F	CYS	76	SG	2.77
1ofr	MN	G	400	G	CYS	76	SG	2.73
1ofr	MN	H	400	B	CYS	76	SG	2.92
1og0	MN	A	400	A	CYS	76	SG	2.77
1og0	MN	B	400	B	CYS	76	SG	2.64
1og0	MN	C	400	C	CYS	76	SG	2.51
1og0	MN	D	400	D	CYS	76	SG	2.50
1og0	MN	E	400	E	CYS	76	SG	2.64
1og0	MN	F	400	F	CYS	76	SG	2.58
1og0	MN	G	400	G	CYS	76	SG	2.55
1og0	MN	H	400	H	CYS	76	SG	2.51
1oqm	MN	B	125	B	MET	344	SD	2.92
1oqm	MN	D	125	D	MET	344	SD	2.87
1u8x	MN	X	700	X	CYS	171	SG	2.29
1vs1	MN	A	301	A	CYS	46	SG	2.58
1vs1	MN	B	301	B	CYS	46	SG	2.66
1vs1	MN	C	301	C	CYS	46	SG	2.74
1vs1	MN	D	301	D	CYS	46	SG	2.53
1yro	MN	B	125	B	MET	344	SD	2.79
1yro	MN	D	125	D	MET	344	SD	2.81
1zco	MN	A	900	A	CYS	31	SG	2.12
2b7o	MN	A	700	A	CYS	87	SG	2.47
2b7o	MN	B	701	B	CYS	87	SG	2.76
2faf	MN	A	1701	A	CYS	307	SG	2.29
2faf	MN	B	1702	B	CYS	307	SG	2.29
2fe6	MN	A	417	A	CYS	357	SG	2.59
2fer	MN	A	417	A	CYS	357	SG	2.51
2feu	MN	A	417	A	CYS	357	SG	2.41
2feu	MN	B	417	B	CYS	357	SG	2.44
2fyd	MN	B	125	B	MET	344	SD	2.79
2fyd	MN	D	527	D	MET	344	SD	2.83
2kfn	MN	_	2	_	S	1008	S	2.61
9ica	MN	A	339	A	STP	338	S1A	1.87

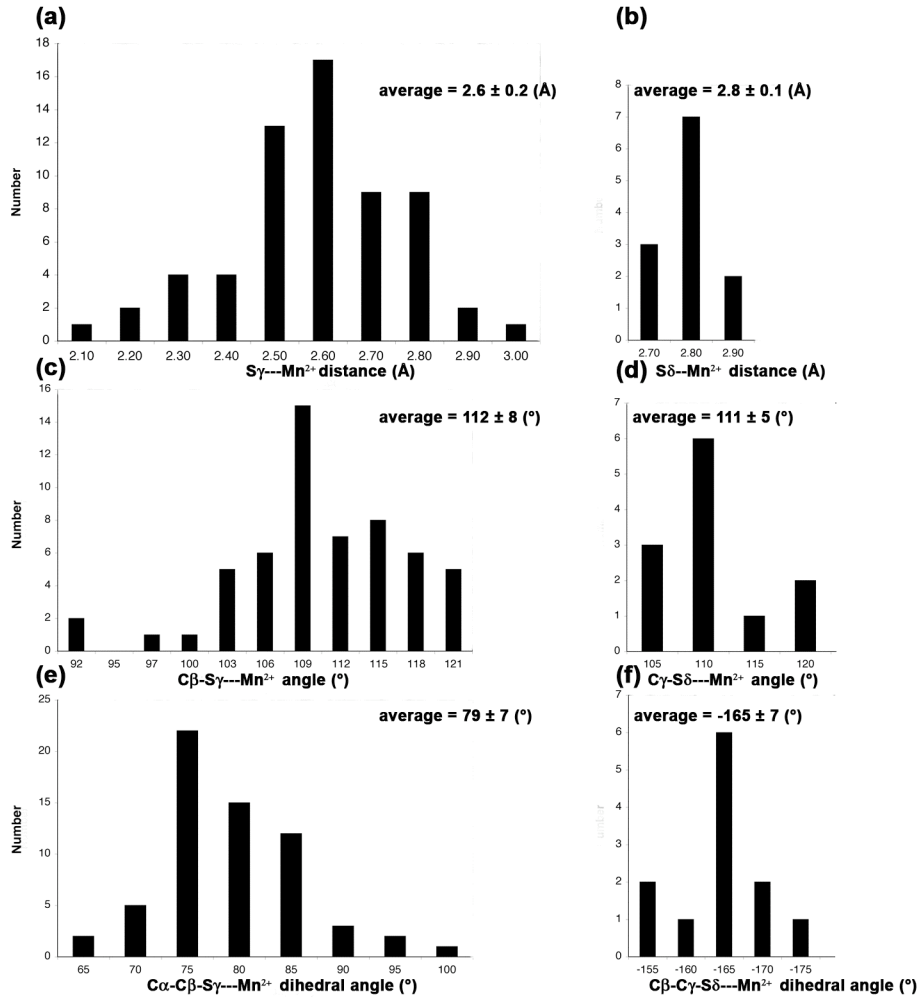

Figure S1: Histograms of S---Mn^{2+} distances for cysteine (a) and methionine (b) coordination; Cβ-Sγ---Mn^{2+} (c) and Cγ-Sδ---Mn^{2+} (d) angles; Cα-Cβ-Sγ---Mn^{2+} (e) and Cβ-Cγ-Sδ---Mn^{2+} (f) dihedral angles.